Modelling household consumption on micro data with a focus on the source of the zeroes*

André Decoster† and Frederic Vermeulen‡

December 16, 2005

Abstract

This paper presents a methodology to estimate household demand on micro data, which takes account of different sources of zero expenditures. It includes both zeroes stemming from abstention and zeroes resulting from infrequency of purchase. The model is applied to the Belgian household budget survey of 1995-1996 for a twelve-commodity breakdown.

KEY WORDS : zero expenditures ; abstention ; infrequency of purchase.

JEL CLASSIFICATION : D12.

1. Introduction

Since Tobin's (1958) seminal paper, considerable attention has been paid to the problem of zero expenditures in household budget surveys. While zero expenditures in the standard univariate tobit model all correspond to corner solutions (consumers may not afford the good at current prices and income), separate processes to generate zeroes are allowed for in bivariate extensions of the tobit

*Financial support of this research by the Fund for Scientific Research - Flanders (contract FWO G.0327.97) and the research fund of the University of Leuven (contract OT 98/03) is gratefully acknowledged. We would like to thank Anton Barten and Bart Capéau for helpful comments on an earlier version of this paper. Any errors remaining are our own.

†Center for Economic Studies, University of Leuven, Naamsestraat 69, B-3000 Leuven, Belgium. E-mail: andre.decoster@econ.kuleuven.be.

‡Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: frederic.vermeulen@uvt.nl.

specification. Cragg's (1971) double hurdle model, for example, not only takes into account zeroes that are corner solutions, but also zeroes that may arise because of abstention. More recently, Deaton and Irish (1984) proposed their P-tobit model which considers yet another source of zeroes, namely those stemming from the combination of infrequency of purchase and the short duration of most household budget surveys. Both models, and extensions of these, were extensively applied in the context of a single-equation analysis (see, for example, Blundell and Meghir, 1987, Jones, 1989, Atkinson, Gomulka and Stern, 1990, Robin, 1993 and Su and Yen, 1996). The approach to model zeroes was quickly extended to a demand system context. Wales and Woodland (1983), Lee and Pitt (1986) and Jarque (1987) proposed modified tobit models to deal with zeroes in demand systems, where all zero expenditures reflect corner solutions. On the other hand, Kay, Keen and Morris (1984), Keen (1986) and Meghir and Robin (1992) take account of an explicit zeroes generating process by considering infrequency of purchase problems in short duration household surveys.

All the above system-based models have one feature in common. They all assume that zeroes arise from the same source, either they reflect corner solutions or they stem from infrequency of purchase. (An exception is Fry and Pashardes, 1994, who model zero expenditures on tobacco, both resulting from abstention and corner solutions, within a demand system. However, zero expenditures on other commodities are ignored). Notwithstanding the fact that this approach is very useful to deal with zeroes on many different commodities, it seems to be a rather implausible assumption once one additionally wants to model commodities that may be subject to abstention. For example, many zero expenditures on tobacco and alcohol originate from abstention, while, simultaneously, zeroes on clothing will be stemming from infrequency of purchase.

In this paper, an empirical model is presented which allows zeroes to result from different sources. More precisely, it takes into consideration zeroes corresponding to abstention and infrequency of purchase. The framework explicitly takes into account that there is preference heterogeneity between consumers and non-consumers of commodities that are subject to abstention (*bads*). It assumes, e.g., that the consumption of tobacco does not 'generate utility' for non-smokers, so that tobacco does not act as an argument in the non-smokers' direct utility function representing their preferences. Consequently, demand functions derived from these preferences do not include the price of tobacco as an explanatory variable. The model thus explicitly assumes that non-smokers are not affected by price changes of tobacco. The framework further allows for the possibility that consumers and non-consumers of bads allocate their budget on the rest of the commodity bundle in a different way. A non-smoker, e.g., might consume more sportswear than a smoker with exactly the same budget on the other commodities, household composition, social status etcetera. Contrary to the a priori assumption of preference heterogeneity with regard to bads, this will mainly be an empirical question in this study.

In summary, the model can be described as follows. Commodities that are possibly subject to abstention are modelled by means of an alternative Cragg specification, which assumes first hurdle dominance (see, e.g., Jones, 1989). This model assumes that positive expenditures on tobacco are in any case observed for smokers, which implies that corner solutions or infrequent tobacco purchases are ruled out. On the other hand, the rest of the commodity bundle is modelled by means of an infrequency of purchase model, conditional on the budget spent on these commodities and on being a consumer or a non-consumer of a bad. This allows for preference heterogeneity with regard to these commodities. Moreover, the approach enables an efficient way of estimating demand in the sense that all observations are used, instead of estimating separate demand systems for consumers and non-consumers of bads. While it has not been done in the model below, corner solutions for these commodities can easily be integrated in the analysis.

The layout of the paper is as follows. In the next section, the methodology is described to model demand with a focus on different sources of zero expenditures. Section 3 gives an empirical illustration for a twelve-commodity breakdown that is based on individual expenditure data drawn from the Belgian household budget survey of 1995-1996. Conclusive remarks are presented in section 4.

2. Modelling zeroes resulting from different sources

2.1. The structure of preferences

The assumption of heterogeneous preferences of consumers and non-consumers of goods which one can abstain from, involves different demand systems for each group of consumers. In fact, two kinds of preference heterogeneity can be discerned[1]. *Preference heterogeneity of the first kind* occurs when some consumers abstain from certain commodities while other consumers do not. This implies different arguments in utility functions (e.g., consumption of alcoholic beverages)

[1]Note that we do not refer here to models that allow for unobserved preference heterogeneity between consumers.

and demand functions (e.g., price of alcoholic beverages) for consumers and non-consumers of bads. Remark that in what follows, the question of which zero expenditures are due to abstention is solved a priori. A *second kind of preference heterogeneity* is that which allows for different preferences of consumers and non-consumers of abstention goods with regard to the rest of the commodity bundle. It may be the case that consumers and non-consumers of a bad allocate their budgets on the other commodities differently. Contrary to preference heterogeneity of the first kind, the approach below permits to test empirically preference heterogeneity of the second kind (see Vermeulen, 2003).

To model both kinds of preference heterogeneity in an efficient way, a specific structure of preferences is needed. To simplify matters somewhat, suppose that only one commodity (tobacco, say) is subject to abstention. Consequently, two different demand systems are required to model commodity demand[2]. Let q_t and $\mathbf{q}_o$ denote the demand on tobacco and the quantity vector of the other commodities respectively. Demographic household characteristics are denoted by the vector $\mathbf{a}$. Consider now the following preference relation :

$$u = D_t \cdot v_S\left(q_t, \mathbf{q}_o; \mathbf{a}\right) + \left(1 - D_t\right) \cdot v_N\left(\mathbf{q}_o; \mathbf{a}\right), \tag{2.1}$$

where $v_S(.)$ and $v_N(.)$ are strictly quasi-concave, increasing and differentiable functions which represent the preferences of smokers and non-smokers respectively. The variable D_t is a discrete shift variable which indicates whether a household contains a smoker ($D_t = 1$) or entirely consists of non-smokers ($D_t = 0$). It is easily seen that the relation in (2.1) explicitly takes into account that tobacco does not play any role in the preferences of non-smokers. This is perfectly plausible, since many non-smokers would not smoke, even if tobacco was available for free. Maximizing the non-smoker's direct utility function $v_N(.)$ subject to his budget constraint $\mathbf{p}_o'\mathbf{q}_o = x$, gives rise to the following Marshallian demand system :

$$\mathbf{q}_o = \mathbf{g}_N\left(x, \mathbf{p}_o; \mathbf{a}\right), \tag{2.2}$$

where x denotes total expenditures and $\mathbf{p}_o$ the price vector of the other commodities. The uncompensated demand of a smoker, on the other hand, equals :

$$\begin{pmatrix} q_t \\ \mathbf{q}_o \end{pmatrix} = \mathbf{g}_S\left(x, p_t, \mathbf{p}_o; \mathbf{a}\right), \tag{2.3}$$

[2] The model gets rapidly involved if households can abstain from many commodities. Due to preference heterogeneity between consumers and non-consumers of the bads, 2^n different demand systems are required if there are n bads.

where, contrary to the non-smoker's case, demand depends on the price of tobacco p_t. Setting aside the problem of zero expenditures on the other commodities, the simplest way to apply the above structure of preferences would be to estimate separate demand systems for smokers and non-smokers. This is a rather inefficient approach. The model can be estimated more parsimoniously if one allows the parameters to vary with D_t. However, under this approach, the non-smoker's demand equations still depend on the price of tobacco. As a consequence, substitution effects between tobacco and the other commodities cannot be avoided. One way to reconcile econometric thrift and separate demand systems (in the sense that different explanatory variables are included in the two systems), is invoking the assumption of weak separability of the set of goods $\mathbf{q}_o$ from tobacco demand q_t for smokers[3]. This changes the preference relation in (2.1) as follows :

$$u = D_t \cdot f_S \left[q_t, v_{So} \left(\mathbf{q}_o ; \mathbf{a} \right) ; \mathbf{a} \right] + \left(1 - D_t \right) \cdot v_N \left(\mathbf{q}_o ; \mathbf{a} \right), \tag{2.4}$$

where $f_S\left(.\right)$ is some increasing function and $v_{So}\left(.\right)$ is a well-behaved subutility function. One immediate consequence of the weakly separable preferences of smokers in (2.4) is that these imply within-group demands $\mathbf{q}_o$ of the following form :

$$\mathbf{q}_o = \mathbf{g}_{So} \left(x^*, \mathbf{p}_o ; \mathbf{a} \right), \tag{2.5}$$

where x^* is the budget spent on the commodity vector $\mathbf{q}_o$. Since the budget x^* equals by definition total expenditures x for non-smokers, a general specification for Marshallian demand for $\mathbf{q}_o$, which encompasses both the demand of smokers (2.5) and non-smokers (2.2), is quickly derived :

$$\mathbf{q}_o = \mathbf{g}_o \left(x^*, \mathbf{p}_o ; \mathbf{a}, D_t \right). \tag{2.6}$$

In equation (2.6), the discrete shift variable D_t acts as a conditioning variable. The above structure of preferences allows to model the within-group demands $\mathbf{q}_o$ without taking account of tobacco. This brings about different demand systems for smokers and non-smokers, that can be estimated in a parsimonious way. A simple test of preference heterogeneity at the level of the commodities $\mathbf{q}_o$ (preference heterogeneity of the second kind) consists of testing whether demand $\mathbf{q}_o$ depends on the conditioning variable D_t (see, Browning and Meghir, 1991). If

[3] The weak separability assumption is not strictly necessary. One can always model the other commodities $\mathbf{q}_o$ via conditional demand functions, whereby demand is conditional on q_t (see, Pollak, 1969 and 1971 and Browning and Meghir, 1991).

the parameters associated with D_t turn out to be insignificant, weak separability of $\mathbf{q}_o$ from D_t holds so that both smokers and non-smokers would have the same within-group demand equations[4]. Demand for tobacco, on the other hand, can be modelled separately, without having to appeal to these within-group demands. This of course because of the weakly separable preferences between tobacco and the other commodities[5].

2.2. The stochastic specification

The discussion of the statistical models that will be used to apply the above structure of preferences consists of two parts. First, we will focus on the modelling of tobacco which is subject to abstention. Second, the infrequency of purchase model to estimate the within-group demands, is discussed. Remark that in what follows, all specifications are in a single-equation context. The reason for this is, that the household budget survey that is available for the empirical application does not contain any price information. Moreover, in Belgium, there is no series of homogeneous budget surveys available that can be used to estimate complete demand systems. Therefore, the estimation of Engel curves that underlie the above structure of preferences, is as far as we can go.

As mentioned in the introduction, we choose to model goods which one can abstain from, by means of an alternative Cragg (1971) specification. This specification is a restricted version of Cragg's double hurdle model, which assumes that two hurdles must be overcome in order to observe positive expenditures on a good (see, for example, Atkinson, Gomulka and Stern, 1990 and Blundell and Meghir, 1987). The first hurdle then captures zeroes coming from abstention, while the second one corresponds to corner solutions. In this paper, we have opted for an alternative that assumes first hurdle dominance (see Jones, 1989). It implies that the participation decision dominates the consumption decision. Consequently, positive expenditures on tobacco will always be observed for smokers, so that corner solutions are ruled out. Given the addictive character of tobacco, this does not

[4]This kind of separability is closely related to Deaton, Ruiz-Castillo and Thomas's (1989) demographic separability. If separability of D_t from $\mathbf{q}_o$ applies, expenditures on tobacco act as a fixed cost for smokers, on top of the expenditures on the other commodities.

[5]Remark that this two-stage budgeting setting is not entirely innocent. In order to allocate total expenditures to the broad commodity groups (tobacco and an aggregate of the other commodities) in an attractive way, one has to resort to an approximation if one wants to avoid rather restrictive functional forms. For a more profound discussion, see Deaton and Muellbauer (1980b).

seem a very bad assumption.

Let q_{th} denote the observed consumption of tobacco of household h and D_{th} household h's discrete shift variable which determines whether the household contains a smoker. The variables z_{th}^* and q_{th}^* are latent variables that correspond to D_{th} and q_{th} and depend on the vectors of household characteristics $\mathbf{a}_{1h}$ and $\mathbf{a}_{2h}$ respectively. The first hurdle dominance model for tobacco can be summarized as follows :

$$z_{th}^* = \mathbf{a}_{1h}'\boldsymbol{\gamma} + u_h, h = 1, ..., H \tag{2.7}$$

$$q_{th}^* = \mathbf{a}_{2h}'\boldsymbol{\beta} + \varepsilon_h, h = 1, ..., H \tag{2.8}$$

$$q_{th} = q_{th}^*, D_{th} = 1 \text{ if } z_{th}^* > 0 \tag{2.9}$$

$$q_{th} = 0, D_{th} = 0 \text{ if } z_{th}^* \leq 0, \tag{2.10}$$

under the assumption of a bivariate normal distribution of the error terms u_h and ε_h and the normalization $\sigma_u^2 = 1$. The sample log-likelihood function of the model takes the form :

$$\log L = \sum_0 \log\left[1 - \Phi\left(\mathbf{a}_{1h}'\boldsymbol{\gamma}\right)\right] + \sum_+ \left[\log \Phi\left(\mathbf{a}_{1h}'\boldsymbol{\gamma}\right) + \log f\left(q_{th} \mid D_{th} = 1\right)\right], \tag{2.11}$$

where $\sum_0$ and $\sum_+$ refer to summation over respectively zero and positive expenditures on tobacco and $f(.)$ denotes a (conditional) normal density function. Remark that equation (2.11) corresponds to Heckman's sample selection model (see, Heckman, 1979). It also consists of the combination of a probit model and a truncated regression model. The model can easily be estimated by means of maximum likelihood which obtains consistent and asymptotically efficient estimators or by means of Heckman's two-step estimator (see, for example, Greene, 1997). Following Pudney (1989), it should be noted that only non-economic household characteristics should appear in equation (2.7). He argues that one abstains from smoking due to personal, social or ethical reasons, rather than economic ones.

Let us now focus on the modelling of the other commodities. One characteristic of short duration household budget surveys is that they contain many zero expenditures, which cannot be attributed to abstention. Zeroes even occur for relatively broad aggregates like clothing, for the simple reason that not every household purchases clothes within the limited period of a budget survey. It is clear that under infrequency of purchase, *consumption* and *purchases* can differ substantially for some goods. Because the data set that will be used in the empirical application contains many zeroes on goods which one usually does not abstain from, we have chosen an infrequency of purchase specification to model the within-group demand (cf. supra). The first formal specification that allows for infrequency of purchase was Deaton and Irish's (1984) P-tobit model, which assumes a constant probability of observing a purchase. A more general model, with different purchase probabilities for different households, was proposed by Blundell and Meghir (1987). The zeroes on the commodities considered by Blundell and Meghir are assumed not to result from corner solutions. Consequently, every modelled commodity is consumed, but a purchase of it is not always observed. We have opted for their approach in what follows. Note however, that this approach can easily be generalized to include the possibility of corner solutions.

In the former section, the problem of zeroes resulting from infrequency of purchase was ignored. However, under these circumstances, *consumption* is no longer observed any more. Indeed, the services of an infrequently purchased commodity are consumed over a period that typically exceeds the duration of the household budget survey. Let q_{oh} denote the *observed purchased quantity* of an infrequently purchased good and D_{oh} a binary indicator that determines whether a purchase is observed ($D_{oh} = 1$) or unobserved ($D_{oh} = 0$). The latent variables are q^*_{oh} (*consumption*) and z^*_{oh}, which respectively correspond with q_{oh} and D_{oh}. As in the first hurdle dominance model, we have two vectors of household characteristics : $\mathbf{a}_{1h}$ and $\mathbf{a}_{2h}$. Purchase behavior (in terms of the probability of observing a purchase of the commodity during the survey) is described by a simple probit model :

$$z^*_{oh} = \mathbf{a}'_{1h}\boldsymbol{\gamma} + u_h, h = 1, ..., H \tag{2.12}$$

$$D_{oh} = 1 \text{ if } z^*_{oh} > 0 \tag{2.13}$$

$$D_{oh} = 0 \text{ if } z^*_{oh} \leq 0, \tag{2.14}$$

where it is assumed that the error term u_h follows a standard normal distribution. The probit model thus implies that $\Pr(D_{oh} = 1) = \Phi(\mathbf{a}'_{1h}\boldsymbol{\gamma})$. As is clear from the model, households may have different purchase probabilities depending on their characteristics. The unobserved consumption, on the other hand, is described as follows :

$$q_{oh}^{*}=\mathbf{a}_{2h}^{\prime}\boldsymbol{\beta}+\varepsilon_{h},h=1,...,H, \tag{2.15}$$

where ε_h is assumed to be normally distributed with zero mean and variance σ_{ε}^{2}. Assuming now that expenditures and consumption are equal on average, $E\left(q_{oh}\right)=E\left(q_{oh}^{*}\right)$, it is easily shown that :

$$E\left(q_{oh}\mid D_{oh}=1\right)\cdot\Pr\left(D_{oh}=1\right)=E\left(q_{oh}^{*}\right). \tag{2.16}$$

Consider, for example, a budget survey which has a duration of one month. If a household buys, on average, one season ticket for public transport every other month, then it follows that the probability of observing the purchase of one season ticket equals one half, while the services drawn from that commodity (consumption) during that month equals half of the ticket. Consequently, observed expenditures will, on average, exceed consumption. The stochastic specification can be completed by filling in the appropriate terms in equation (2.16) and introducing an error term υ_h :

$$\Phi\left(\mathbf{a}_{1h}^{\prime}\boldsymbol{\gamma}\right)\cdot q_{oh}=q_{oh}^{*}+\upsilon_{h}=\left(\mathbf{a}_{2h}^{\prime}\boldsymbol{\beta}+\varepsilon_{h}\right)+\upsilon_{h}=\mathbf{a}_{2h}^{\prime}\boldsymbol{\beta}+\eta_{h}\text{ for }q_{oh}>0, \tag{2.17}$$

where $\eta_h=\varepsilon_h+\upsilon_h$. Note that independence is assumed between the purchase behavior and consumption in equation (2.17). More precisely, this implies independence between the composite error term η_h of equation (2.17) and the probit error term u_h of equation (2.12). The sample log-likelihood of this infrequency of purchase model now takes the following form :

$$\begin{aligned}\log L &= \sum_{0}\log\left[1-\Phi\left(\mathbf{a}_{1h}^{\prime}\boldsymbol{\gamma}\right)\right] \qquad (2.18)\\ &\quad+\sum_{+}\left[-\log\sigma_{\eta}+\log\phi\left(\left(\Phi\left(\mathbf{a}_{1h}^{\prime}\boldsymbol{\gamma}\right)\cdot q_{oh}-\mathbf{a}_{2h}^{\prime}\boldsymbol{\beta}\right)/\sigma_{\eta}\right)+2\log\Phi\left(\mathbf{a}_{1h}^{\prime}\boldsymbol{\gamma}\right)\right].\end{aligned}$$

Note that the second $\log\Phi\left(\mathbf{a}_{1h}^{\prime}\boldsymbol{\gamma}\right)$ in the final part of equation (2.18) corresponds to the Jacobian $\left|\frac{\partial q_{oh}^{*}}{\partial q_{oh}}\right|$, due to the change from purchases to unobservable consumption. This infrequency of purchase model can be estimated consistently and (asymptotically) efficiently via maximum likelihood.

Some final remarks have to be made. First, note that equation (2.18) provides an adequate instrument to test for the interesting benchmark of Deaton and Irish's (1984) P-tobit specification. Indeed, if one assumes that the vector $\mathbf{a}_{1h}$ only contains a constant term, then the probability of observing positive expenditures on a commodity is independent of household characteristics. Consequently, every household will purchase the commodity with the same probability. The null hypothesis of having a P-tobit model can easily be tested by means of a likelihood ratio test. Second, remark that Engel curves derived from equation (2.18) will not automatically satisfy the adding-up restriction, like they do when estimated by means of ordinary least squares. Therefore, one commodity in the within-group demand system will be treated as a residual with no specific functional form of its own. The spending on that commodity is then the difference between the within-group budget and the expenditures on the directly modelled demands of this within-group system. Note that the choice of that residual commodity may affect the obtained results. Linear Engel curves (Keen, 1986) or the introduction of more information on purchase behavior (Meghir and Robin, 1992 and Robin, 1993) would avoid this residual commodity. However, linear Engel curves are very restrictive and information on the number of purchases is lacking in the household budget survey that will be used in the next section.

3. Empirical illustration

3.1. Overview

Both the first hurdle dominance and the infrequency of purchase model are applied within the structure of preferences derived in the former section. This is done on the Belgian household budget survey of 1995-1996. The survey consists of expenditures on over 800 commodities for a sample of 2724 households. Each household had to register all its expenditures on these commodities during one month. The survey further provides about thirty socio-economic household and personal characteristics like family size, occupation, age and social status of the family members. Price information is lacking. Moreover, a time series of homogeneous household budget surveys is unavailable in Belgium. Consequently, complete systems of demand equations cannot easily be estimated on Belgian micro data. Therefore, we concentrate on Engel curves in this empirical application.

We will model a twelve-commodity breakdown within the structure of preferences of the former section. One of the goods, tobacco, is subject to abstention and is estimated by means of the first hurdle dominance model. The other commodities of the system are : food, beverages, clothing, rent, energy, durables, maintenance, personal care, transportation, services and leisure activities. Food and rent are estimated via simple OLS, since there are no zero expenditures on these commodities. Leisure is treated as a residual commodity in order to satisfy adding-up. The remaining goods are estimated via the infrequency of purchase model outlined in the former section.

The Engel curves derived from the Quadratic Almost Ideal Demand System (QUAIDS) of Banks, Blundell and Lewbel (1997) are chosen as functional form for all modelled commodities. This rank three demand system is a quadratic extension of Deaton and Muellbauer's (1980a) Almost Ideal Demand System and has the nice property that it permits goods to be luxuries at low income levels, while they can become necessities if the household is getting better off. The Engel curves of QUAIDS are as follows :

$$w_{ih} = \alpha_{ih} + \beta_{ih} \log m_h + \lambda_{ih} (\log m_h)^2, \; i = 1, ..., n \text{ and } h = 1, ..., H, \tag{3.1}$$

where w_{ih} is household h's budget share of commodity i and m_h is the budget spent on the n commodities. Note that m_h equals total expenditures x_h in the case of the first-stage allocation (tobacco), and x_h^* with respect to the within-group demands (the other commodities). Further, the coefficients α_{ih}, β_{ih} and λ_{ih} may depend on some household characteristics (see Appendix for a description of the variables taken up in the analysis). One important case which should be considered in our setting, is the dependence of these coefficients on the discrete shift variable D_{th}, which determines whether or not the household contains a smoker. If these coefficients are dependent on D_{th}, then the null hypothesis of no preference heterogeneity of the second kind between smokers and non-smokers can be rejected.

3.2. First hurdle dominance model for tobacco expenditures

The maximum likelihood parameter estimates of equation (2.11) are presented in table 1, together with their standard errors. The second and the third column of the table show the results of the participation decision on smoking. Note that following Pudney's (1989) argumentation, economic variables like total expenditures x are not taken up in this first hurdle. Although the model seems to be a bit overparameterized, the results are fairly plausible. Households with a blue collar (S_1) or unemployed (S_4) head of the family are more likely to contain a smoker than the benchmark case of households of which the head of the family is self-employed (all social status variables equal to zero). On the other hand, the variable S_3, which captures households with a retired head of the family, has a significant negative impact on the participation decision. The other explanatory variables have insignificant effects.

As to the Engel curve maximum likelihood estimates (last two columns of table 1), the standard errors are rather large for most of the independent variables. There seems to be no significant relation between the logarithm of total expenditures (or other variables where expenditures are involved) and the budget share of tobacco, once one has decided to smoke. The consumption of tobacco is significantly higher for households living in the Walloon Region (G_2) and for households with a blue collar (S_1) or an unemployed (S_4) head of the family. Consumption is significantly lower if the head of the family is retired (S_3). Note that the null hypothesis of no sample selection bias can be rejected, since the correlation coefficient between the errors u_h and ε_h of the participation decision and the consumption equation is highly significant. Consequently, the estimation of an Engel curve for tobacco on the subsample of smokers, would have obtained inconsistent estimates.

	Table 1 : Tobacco			
	Participation decision		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-0.283	0.189	0.331	0.885
$\log x$			-0.049	0.137
$(\log x)^2$			0.002	0.005
$\log x \cdot V$			-0.001	0.005
$\log x \cdot K_1$			0.000	0.003
$\log x \cdot K_2$			-0.001	0.007
$\log x \cdot Age$			-0.000	0.002
G_1	0.059	0.176	0.004	0.006
G_2	0.128	0.066	0.007	0.003
S_1	0.314	0.117	0.016	0.004
S_2	0.059	0.113	0.004	0.004
S_3	-0.303	0.153	-0.013	0.006
S_4	0.474	0.165	0.021	0.007
Occ	-0.084	0.069	-0.004	0.003
Age	-0.044	0.044	-0.002	0.028
$Dens$	0.000	0.000	-0.000	0.000
V	0.021	0.049	0.011	0.063
K_1	-0.012	0.043	-0.006	0.453
K_2	-0.063	0.073	0.017	0.097
σ_ε			0.047	0.001
$\rho_{\varepsilon u}$			0.999	0.000
$logL$: 768.2632				

3.3. Infrequency of purchase model for within-group demand

The estimation results of the within-group Engel curves of food, beverages, clothing, rent, energy, durables, maintenance, personal care, transportation and services are presented in the Appendix. First, note that no purchase probability is estimated for the commodities food and rent, for the simple reason that there are no zero expenditures on these commodities. In order to make the model not too complex, it is assumed that both smokers and non-smokers have the same purchase probabilities. Contrary to the first hurdle in the first hurdle dominance model, economic variables can occur in this purchase probability. Therefore, group expenditures x^* is taken up in the latter. The sign of its effect is an empirical question.

Richer households, for example, more often buy clothes, but may be observed to buy beverages fewer times due to the fact that they buy larger quantities of it. Further, we assumed that there may be preference heterogeneity between smokers and non-smokers with regard to the intercept and the coefficients associated with the explanatory variables that involve x^*.

Instead of discussing the estimations in a very detailed way, we will restrict ourselves to some particular results. Although some purchase probabilities and Engel curves appear overparameterized, the infrequency of purchase model seems to be plausible in general. All purchase probabilities and most of the Engel curves depend significantly on group expenditures x^*. Many Engel curves also have significant expenditure interaction terms. Interesting to mention is that the coefficient associated with the quadratic expenditure term $(\log x^*)^2$ is significant for 4 commodities : clothing, rent, maintenance and transportation. This seems to justify the quadratic extension of the Almost Ideal Demand System.

Let us now focus on the question whether there is preference heterogeneity of the second kind between smokers and non-smokers. This is easily tested by means of a likelihood ratio test. As is clear from table 2, the null hypothesis of no preference heterogeneity is rejected for 4 commodities at a 5% significance level. Although within-group demand is assumed to be separable from tobacco expenditure, smokers allocate the budget on the other commodities in another way than non-smokers do.

Table 2 : Test of preference heterogeneity

Commodity	Test statistic	χ^2_7	Conclusion
Food	19.231	14.067	H_0 rejected
Beverages	38.870	14.067	H_0 rejected
Clothing	16.034	14.067	H_0 rejected
Rent	40.252	14.067	H_0 rejected
Energy	14.006	14.067	H_0 not rejected
Durable goods	9.872	14.067	H_0 not rejected
Maintenance	4.046	14.067	H_0 not rejected
Personal care	5.240	14.067	H_0 not rejected
Transportation	10.598	14.067	H_0 not rejected
Services	11.316	14.067	H_0 not rejected

Finally, table 3 shows the likelihood ratio test results of Deaton and Irish's (1984) more restrictive P-tobit specification. The null hypothesis of having a P-tobit is strongly rejected for all modelled within-group commodities. Consequently, household characteristics indeed have a role to play in the determination of the purchase probabilities.

Table 3 : Test of P-tobit

Commodity	Test statistic	χ^2_{13}	Conclusion
Beverages	195.550	22.362	H_0 rejected
Clothing	195.850	22.362	H_0 rejected
Energy	342.914	22.362	H_0 rejected
Durable goods	204.196	22.362	H_0 rejected
Maintenance	395.642	22.362	H_0 rejected
Personal care	154.454	22.362	H_0 rejected
Transportation	418.404	22.362	H_0 rejected
Services	49.494	22.362	H_0 rejected

3.4. Total expenditure elasticities

We conclude this section with presenting the total expenditure elasticities of the twelve-commodity breakdown[6]. These expenditure elasticities were calculated for both smokers and non-smokers and were evaluated at average budget shares and average expenditures of those households consuming the good[7]. As is clear from table 4, the response to income changes is very similar for both groups of consumers. Food, beverages, rent and energy are necessities for both groups, while clothing, durable goods, maintenance, transportation, services and leisure activities are luxuries. The only exception is personal care, although the difference between the elasticities is not very pronounced. Also for the rest of the commodities, the magnitude of the elasticities of smokers and non-smokers does not differ much.

[6]Note that the coefficients of the demand for leisure activities are indirectly calculated via the adding-up restriction.

[7]Remark that the *total* expenditure elasticities of smokers are obtained by multiplying the *first-stage* expenditure elasticitities (allocation of total expenditures to tobacco and an aggregate of the other commodities) by the *second-stage* expenditure elasticities (allocation of group expenditures to within-group demands), see Edgerton (1997).

Table 4 : Total expenditure elasticities

Commodity	Non-smokers	Smokers
Tobacco	-	0.919
Food	0.565	0.580
Beverages	0.966	0.810
Clothing	1.095	1.313
Rent	0.279	0.341
Energy	0.585	0.450
Durable goods	2.500	2.421
Maintenance	1.173	1.182
Personal care	1.030	0.976
Transportation	1.113	1.063
Services	1.041	1.060
Leisure activities	1.787	3.095

4. Conclusion

This paper presents a methodology to model household demand on micro data, which takes account of different sources of zero expenditures. It not only considers zeroes stemming from abstention, but also zeroes resulting from infrequency of purchase. Commodities which are possibly subject to abstention (tobacco, for example) are modelled by means of a first hurdle dominance model. The rest of the commodity bundle is modelled by an infrequency of purchase specification, conditional on being a consumer or a non-consumer of the commodities which one can abstain from. Consequently, different demand systems are obtained for different groups of consumers. The framework is applied for a twelve-commodity breakdown on the Belgian household budget survey of 1995-1996.

References

[1] Atkinson, A.B., J. Gomulka and N.H. Stern (1990), "Spending on alcohol : evidence from the Family Expenditure Survey 1970-1983", *Economic Journal*, 100, 808-827.

[2] Banks, J., R. Blundell and A. Lewbel (1997), "Quadratic Engel curves and consumer demand", *Review of Economics and Statistics*, 74, 527-539.

[3] Blundell, R. and C. Meghir (1987), "Bivariate alternatives to the tobit model", *Journal of Econometrics*, 34, 179-200.

[4] Browning, M. and C. Meghir (1991), "The effects of male and female labor supply on commodity demands", *Econometrica*, 59, 925-951.

[5] Cragg, J.G. (1971), "Some statistical models for limited dependent variables with application to the demand for durable goods", *Econometrica*, 39, 829-844.

[6] Deaton, A. and M. Irish (1984), "Statistical models for zero expenditures in household budgets", *Journal of Public Economics*, 23, 59-80.

[7] Deaton, A. and J. Muellbauer (1980a), "An almost ideal demand system", *American Economic Review*, 70, 312-326.

[8] Deaton, A. and J. Muellbauer (1980b), *Economics and consumer behavior*, Cambridge, Cambridge University Press.

[9] Deaton, A., J. Ruiz-Castillo and D. Thomas (1989), "The influence of household composition on household expenditure patterns : theory and Spanish evidence", *Journal of Political Economy*, 97, 179-200.

[10] Edgerton, D.L. (1997), "Weak separability and the estimation of elasticities in multistage demand systems", *American Journal of Agricultural Economics*, 79, 62-79.

[11] Fry, V. and P. Pashardes (1994), "Abstention and aggregation in consumer demand : zero tobacco expenditures", *Oxford Economic Papers*, 46, 502-518.

[12] Greene, W.H. (1997), *Econometric analysis*, New Jersey : Prentice Hall.

[13] Heckman, J.J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47, 153-161.

[14] Jarque, C.M. (1987), "An application of LDV models to household expenditure analysis in Mexico", *Journal of Econometrics*, 36, 31-53.

[15] Jones, A.M. (1989), "A double-hurdle model of cigarette consumption", *Journal of Applied Econometrics*, 4, 23-39.

[16] Kay, J.A., M.J. Keen and C.N. Morris (1984), "Estimating consumption from expenditure data", *Journal of Public Economics*, 23, 169-181.

[17] Keen, M.J. (1986), "Zero expenditures and the estimation of Engel curves", *Journal of Applied Econometrics*, 1, 277-286.

[18] Lee, L.-F. and M.M. Pitt (1986), "Microeconometric demand systems with binding nonnegativity constraints : the dual approach", *Econometrica*, 54, 1237-1242.

[19] Meghir, C. and J.-M. Robin (1992), "Frequency of purchase and the estimation of demand systems", *Journal of Econometrics*, 53, 53-85.

[20] Pollak, R.A. (1969), "Conditional demand functions and consumption theory", *Quarterly Journal of Economics*, 83, 60-78.

[21] Pollak, R.A. (1971), "Conditional demand functions and the implications of separable utility", *Southern Economic Journal*, 37, 423-433.

[22] Pudney, S. (1989), *Modelling individual choice. The econometrics of corners, kinks and holes*, Oxford : Basil Blackwell.

[23] Robin, J.-M. (1993), "Econometric analysis of the short-run fluctuations of households' purchases", *Review of Economic Studies*, 60, 923-934.

[24] Tobin, J. (1958), "Estimation of relationships for limited dependent variables", *Econometrica*, 26, 24-36.

[25] Su, S.-J.B. and S.T. Yen (1996), "Microeconometric models of infrequently purchased goods : an application to household pork consumption", *Empirical Economics*, 21, 513-533.

[26] Vermeulen, F. (2003), "Do smokers behave differently? A tale of zero expenditures and separability concepts", *Economics Bulletin*, 4, 1-7.

[27] Wales, T.J. and A.D. Woodland (1983), "Estimation of consumer demand systems with binding non-negativity constraints", *Journal of Econometrics*, 21, 263-285.

A. Appendix

Table A1 : Glossary of variables in the analysis

		Mean
x	Total expenditures (annual amount in BEF)	999,668
x^*	Budget spent on other commodities (tobacco excluded)	989,702
D_t	Dummy variable for smoker, = 1 if smoker	0.373
V	Number of adults	1.919
K_1	Number of children $\geq$ 6 years of age	0.399
K_2	Number of children < 6 years of age	0.198
Age	Age class of head of the family (4 classes of 10 years, fifth class $\geq$ 60)	2.364
G_1	Dummy region, = 1 Brussels Region (Flemish Region is benchmark)	0.109
G_2	Dummy region, = 1 Walloon Region (Flemish Region is benchmark)	0.326
S_1	Dummy social status head of the family, = 1 if blue collar (self-empl. bench.)	0.193
S_2	Dummy social status head of the family, = 1 if white collar (self-empl. bench.)	0.294
S_3	Dummy social status head of the family, = 1 if retired (self-empl. bench.)	0.321
S_4	Dummy social status head of the family, = 1 if unoccupied (self-empl. bench.)	0.089
Occ	Number of persons occupied	0.947
$Dens$	Population density municipality (per squared kilometer)	1765

Table A2 : Percentage zero expenditures

	%
Tobacco	62.72
Food	0.00
Beverages	2.23
Clothing	4.05
Rent	0.00
Energy	1.24
Durables	5.16
Maintenance	3.76
Personal care	2.15
Transportation	6.87
Leisure	0.16
Services	0.26

Table A3 : Food (OLS estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$			-0.775	0.834
D_t			1.998	1.048
$\log x^*$			0.127	0.124
$\log x^* \cdot D_t$			-0.290	0.154
$(\log x^*)^2$			-0.005	0.005
$(\log x^*)^2 \cdot D_t$			0.011	0.006
$\log x^* \cdot V$			-0.200	0.003
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			-0.005	0.003
$\log x^* \cdot K_1 \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_2$			-0.010	0.005
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			-0.006	0.002
$\log x^* \cdot Age \cdot D_t$			0.000	0.000
G_1			-0.002	0.005
G_2			0.010	0.002
S_1			0.008	0.004
S_2			-0.003	0.004
S_3			-0.003	0.005
S_4			0.006	0.006
Occ			-0.001	0.002
Age			0.085	0.025
$Dens$			0.000	0.000
V			0.310	0.044
K_1			0.093	0.046
K_2			0.151	0.074
σ_η			0.060	0.000
$logL$: 4121.037				

Table A4 : Beverages (ML estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-8.557	1.590	-0.462	0.435
D_t			-0.094	0.542
$\log x^*$	0.800	0.120	0.068	0.064
$\log x^* \cdot D_t$			0.019	0.080
$(\log x^*)^2$			-0.002	0.002
$(\log x^*)^2 \cdot D_t$			-0.001	0.003
$\log x^* \cdot V$			-0.001	0.002
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			0.001	0.002
$\log x^* \cdot K_1 \cdot D_t$			0.000	0.000
$\log x^* \cdot K_2$			-0.002	0.003
$\log x^* \cdot K_2 \cdot D_t$			0.000	0.000
$\log x^* \cdot Age$			-0.000	0.001
$\log x^* \cdot Age \cdot D_t$			0.000	0.000
G_1	-0.730	0.181	0.003	0.003
G_2	-0.147	0.109	0.005	0.001
S_1	0.229	0.278	0.001	0.002
S_2	0.276	0.237	-0.001	0.002
S_3	-0.183	0.366	-0.000	0.003
S_4	-0.144	0.327	-0.002	0.003
Occ	0.016	0.171	0.000	0.001
Age	-0.099	0.085	0.006	0.013
$Dens$	-0.000	0.000	0.000	0.000
V	0.077	0.077	0.016	0.021
K_1	0.361	0.238	-0.010	0.025
K_2	-0.108	0.166	0.021	0.046
σ_η			0.021	0.000
$logL$: 8537.144				

Table A5 : Clothing (ML estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-13.299	1.700	-1.707	0.801
D_t			2.537	0.997
$\log x^*$	1.114	0.129	0.260	0.118
$\log x^* \cdot D_t$			-0.380	0.145
$(\log x^*)^2$			-0.010	0.004
$(\log x^*)^2 \cdot D_t$			0.014	0.005
$\log x^* \cdot V$			0.006	0.002
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			0.004	0.003
$\log x^* \cdot K_1 \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_2$			-0.001	0.004
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			-0.001	0.002
$\log x^* \cdot Age \cdot D_t$			0.000	0.000
G_1	-0.120	0.228	-0.008	0.004
G_2	-0.150	0.103	-0.012	0.002
S_1	0.425	0.251	0.007	0.003
S_2	0.141	0.222	0.004	0.003
S_3	0.166	0.262	-0.000	0.005
S_4	0.264	0.275	-0.002	0.005
Occ	0.010	0.183	0.005	0.002
Age	-0.051	0.067	0.015	0.022
$Dens$	-0.000	0.000	-0.000	0.000
V	0.078	0.093	-0.081	0.035
K_1	0.231	0.157	-0.044	0.036
K_2	-0.116	0.152	0.020	0.062
σ_η			0.037	0.000
$logL$: 6857.776				

Table A6 : Rent (OLS estimation)

Variable	Purchase probability: Parameter estimate	Purchase probability: Standard error	Engel curve estimates: Parameter estimate	Engel curve estimates: Standard error
$Intercept$			-3.302	1.261
D_t			-0.020	1.583
$\log x^*$			0.671	0.188
$\log x^* \cdot D_t$			-0.015	0.233
$(\log x^*)^2$			-0.030	0.007
$(\log x^*)^2 \cdot D_t$			0.001	0.009
$\log x^* \cdot V$			0.017	0.005
$\log x^* \cdot V \cdot D_t$			0.001	0.000
$\log x^* \cdot K_1$			0.006	0.005
$\log x^* \cdot K_1 \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_2$			0.008	0.008
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			0.005	0.003
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1			0.044	0.008
G_2			-0.003	0.003
S_1			-0.024	0.006
S_2			-0.019	0.005
S_3			-0.007	0.008
S_4			-0.014	0.008
Occ			0.004	0.004
Age			-0.057	0.037
$Dens$			-0.000	0.000
V			-0.239	0.067
K_1			-0.087	0.069
K_2			-0.106	0.112
σ_η			0.094	0.000
$logL$: 2996.575				

Table A7 : Energy (ML estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-18.346	1.341	-0.737	0.407
D_t			-0.167	0.479
$\log x^*$	1.490	0.101	0.133	0.061
$\log x^* \cdot D_t$			0.029	0.071
$(\log x^*)^2$			-0.006	0.002
$(\log x^*)^2 \cdot D_t$			-0.001	0.003
$\log x^* \cdot V$			-0.001	0.002
$\log x^* \cdot V \cdot D_t$			0.000	0.000
$\log x^* \cdot K_1$			-0.003	0.002
$\log x^* \cdot K_1 \cdot D_t$			0.000	0.000
$\log x^* \cdot K_2$			-0.001	0.003
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			0.002	0.001
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1	-0.095	0.181	-0.003	0.003
G_2	-0.047	0.056	0.003	0.001
S_1	0.432	0.170	0.001	0.002
S_2	0.480	0.174	0.000	0.002
S_3	-0.566	0.196	0.004	0.003
S_4	-0.424	0.198	0.002	0.003
Occ	-0.335	0.089	0.002	0.001
Age	0.085	0.040	-0.017	0.015
$Dens$	0.000	0.000	-0.000	0.000
V	0.104	0.098	0.022	0.024
K_1	-0.029	0.074	0.049	0.025
K_2	-0.303	0.075	0.017	0.048
σ_η			0.023	0.000
$logL$: 8470.168				

Table A8 : Durable goods (ML estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-16.373	1.813	12.091	1.523
D_t			0.135	2.11
$\log x^*$	1.367	0.139	-1.942	0.225
$\log x^* \cdot D_t$			-0.011	0.304
$(\log x^*)^2$			0.078	0.008
$(\log x^*)^2 \cdot D_t$			0.000	0.011
$\log x^* \cdot V$			-0.015	0.006
$\log x^* \cdot V \cdot D_t$			-0.001	0.000
$\log x^* \cdot K_1$			-0.025	0.006
$\log x^* \cdot K_1 \cdot D_t$			0.000	0.000
$\log x^* \cdot K_2$			-0.021	0.009
$\log x^* \cdot K_2 \cdot D_t$			-0.001	0.001
$\log x^* \cdot Age$			-0.016	0.003
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1	-0.005	0.248	-0.019	0.009
G_2	-0.054	0.105	-0.011	0.004
S_1	0.341	0.255	0.018	0.006
S_2	0.209	0.258	0.014	0.005
S_3	0.163	0.378	0.004	0.009
S_4	0.247	0.365	0.012	0.009
Occ	0.073	0.182	-0.017	0.004
Age	-0.058	0.077	0.212	0.046
$Dens$	-0.000	0.000	-0.000	0.000
V	-0.062	0.087	0.195	0.088
K_1	0.089	0.103	0.328	0.088
K_2	0.302	0.313	0.287	0.127
σ_η			0.082	0.001
$logL$: 4698.216				

	Table A9 : Maintenance (ML estimation)			
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	5.807	0.690	0.935	0.220
D_t			-0.203	0.365
$\log x^*$	-0.425	0.052	-0.133	0.033
$\log x^* \cdot D_t$			0.029	0.053
$(\log x^*)^2$			0.005	0.001
$(\log x^*)^2 \cdot D_t$			-0.001	0.002
$\log x^* \cdot V$			0.000	0.001
$\log x^* \cdot V \cdot D_t$			0.000	0.000
$\log x^* \cdot K_1$			-0.000	0.001
$\log x^* \cdot K_1 \cdot D_t$			0.000	0.000
$\log x^* \cdot K_2$			0.001	0.002
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			0.002	0.001
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1	-0.353	0.078	-0.000	0.002
G_2	-0.059	0.036	0.000	0.001
S_1	0.299	0.133	-0.003	0.002
S_2	0.727	0.144	-0.001	0.001
S_3	-0.909	0.230	-0.001	0.002
S_4	-0.317	0.204	-0.002	0.002
Occ	-0.528	0.111	0.000	0.001
Age	-0.001	0.045	-0.020	0.009
$Dens$	0.000	0.000	-0.000	0.000
V	1.146	0.058	-0.002	0.015
K_1	0.524	0.091	0.003	0.014
K_2	0.078	0.100	-0.005	0.023
σ_η			0.015	0.000
$logL$: 8832.114				

Table A10 : Personal care (ML estimation)				
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-13.620	2.256	-2.432	0.949
D_t			-1.165	1.388
$\log x^*$	1.099	0.178	0.365	0.141
$\log x^* \cdot D_t$			0.174	0.204
$(\log x^*)^2$			-0.013	0.005
$(\log x^*)^2 \cdot D_t$			-0.006	0.008
$\log x^* \cdot V$			0.006	0.004
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			0.002	0.004
$\log x^* \cdot K_1 \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_2$			-0.002	0.005
$\log x^* \cdot K_2 \cdot D_t$			-0.000	0.000
$\log x^* \cdot Age$			-0.002	0.002
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1	0.052	0.323	0.001	0.005
G_2	0.219	0.127	0.005	0.002
S_1	0.604	0.273	0.009	0.005
S_2	0.331	0.206	0.007	0.005
S_3	0.391	0.427	0.029	0.006
S_4	0.724	0.368	0.023	0.006
Occ	0.239	0.233	-0.001	0.003
Age	-0.008	0.082	0.026	0.027
$Dens$	0.000	0.000	0.000	0.000
V	0.101	0.109	-0.087	0.049
K_1	-0.123	0.121	-0.036	0.054
K_2	0.231	0.344	0.045	0.073
σ_η			0.046	0.001
$logL$: 6583.929				

Table A11 : Transportation (ML estimation)

Variable	Purchase probability: Parameter estimate	Purchase probability: Standard error	Engel curve estimates: Parameter estimate	Engel curve estimates: Standard error
$Intercept$	-15.163	1.511	-3.766	0.858
D_t			0.390	1.177
$\log x^*$	1.205	0.117	0.561	0.126
$\log x^* \cdot D_t$			-0.054	0.172
$(\log x^*)^2$			-0.020	0.005
$(\log x^*)^2 \cdot D_t$			0.002	0.006
$\log x^* \cdot V$			-0.001	0.003
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			0.009	0.003
$\log x^* \cdot K_1 \cdot D_t$			0.000	0.000
$\log x^* \cdot K_2$			0.007	0.005
$\log x^* \cdot K_2 \cdot D_t$			0.001	0.000
$\log x^* \cdot Age$			0.003	0.002
$\log x^* \cdot Age \cdot D_t$			0.000	0.000
G_1	-0.117	0.200	-0.001	0.004
G_2	0.013	0.079	0.008	0.002
S_1	-0.081	0.173	0.003	0.003
S_2	0.495	0.217	0.009	0.003
S_3	-0.348	0.251	-0.007	0.005
S_4	0.028	0.243	-0.001	0.005
Occ	-0.127	0.125	-0.001	0.002
Age	0.047	0.048	-0.052	0.025
$Dens$	0.000	0.000	0.000	0.000
V	0.277	0.080	0.017	0.039
K_1	0.234	0.099	-0.123	0.037
K_2	-0.043	0.093	-0.110	0.065
σ_η			0.038	0.000
$logL$: 6380.427				

	Table A12 : Services (ML estimation)			
	Purchase probability		Engel curve estimates	
Variable	Parameter estimate	Standard error	Parameter estimate	Standard error
$Intercept$	-28.163	9.053	-0.437	0.787
D_t			-1.701	1.196
$\log x^*$	2.421	0.728	0.086	0.116
$\log x^* \cdot D_t$			0.247	0.175
$(\log x^*)^2$			-0.003	0.004
$(\log x^*)^2 \cdot D_t$			-0.009	0.006
$\log x^* \cdot V$			-0.001	0.003
$\log x^* \cdot V \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_1$			-0.003	0.004
$\log x^* \cdot K_1 \cdot D_t$			-0.000	0.000
$\log x^* \cdot K_2$			0.017	0.005
$\log x^* \cdot K_2 \cdot D_t$			0.000	0.000
$\log x^* \cdot Age$			0.004	0.002
$\log x^* \cdot Age \cdot D_t$			-0.000	0.000
G_1	0.197	1.344	0.009	0.005
G_2	0.249	0.990	0.004	0.002
S_1			-0.020	0.004
S_2			-0.014	0.003
S_3			-0.023	0.005
S_4			-0.024	0.006
Occ			0.002	0.003
Age	0.131	0.248	-0.061	0.027
$Dens$	-0.000	0.000	-0.000	0.000
V	-0.189	1.497	0.011	0.046
K_1	-0.360	1.261	0.043	0.056
K_2	-0.171	3.066	-0.234	0.073
σ_η			0.051	0.000
$logL$: 6681.542				

Convergence problems arose for services when the same variables as for the other commodities were taken up in the purchase probability. Skipping the social status variables ($S_1 - S_4$) and the number of occupied household members (Occ) solved this problem.